Mention-some readings of plural-marked questions: Experimental evidence*

Yimei Xiang\textsuperscript{1} & Alexandre Cremers\textsuperscript{2}

\textsuperscript{1}Harvard University, \textsuperscript{2}ILLC, Universiteit van Amsterdam

1. What is mention-some?

In most daily conversations, a question admits only an exhaustive answer. For example, to properly answer (1), the addressee needs to specify all the attendants to the party, as in (1a). If the addressee doesn’t have enough knowledge of this question and can only provide a non-exhaustive answer, then she needs to indicate the incompleteness of her answer. For instance, she can mark her answer with a prosodic rise-fall-rise contour (indicated by ‘.../’), as in (1b). We call (1a) a complete answer while (1b) a partial answer. If a partial answer is not properly marked, such as taking a falling tone as in (1c) (indicated by ‘\’), it gives rise to an undesired exhaustivity inference.

(1) Who went to the party?
\hspace{1cm} (w: Only John and Mary went to the party.)
\hspace{1cm} a. John and Mary.
\hspace{1cm} b. John did .../ \hspace{1cm} \sim I don’t know who else did.
\hspace{1cm} c. #John did.\hspace{1cm} \sim Only John did.

But, \diamond -questions (i.e., questions with a possibility modal) admit not only exhaustive answers but also non-exhaustive answers (Groenendijk & Stokhof 1984). For instance, (2) can be naturally addressed by specifying one or all of the accessible coffee stores, as in (2a) and (2b)/(2c) respectively. Moreover, while being non-exhaustive, the answer (2a) does not need to carry an ignorance mark. In other words, (2a) does not yield an exhaustivity inference even if taking a falling tone. We call answers like (2a) mention-some answers, and answers like (2b)/(2c) mention-all answers. Questions admitting mention-some answers are called mention-some questions.

*We wish to thank audiences at NELS 47 and the Inquisitive Semantics seminar at ILLC, University van Amsterdam. The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.313610 and was supported by ANR-10-IDEX-0001-02 PSL* and ANR-10-LABX-0087 IEC.
Where can we get coffee?

(w: There are only two accessible coffee stores: A and B)

a. Store A. \(\rightarrow\) We can get coffee at Store A only.
b. Store A and Store B.
c. Store A or Store B.

Crucially, mention-some is a very special form of non-exhaustivity: a mention-some answer specifies exactly one of the possible options \textbf{[Xiang 2016b: ch. 2]}. In replying to a \(\Diamond\)-question, if an answer provides multiple choices and is not ignorance-marked, it is interpreted exhaustively, as shown in \textbf{[3b]} Hence, mention-some is more precisely \textit{mention-one}.

Who can chair the committee?

a. Andy. \(\rightarrow\) Only Andy can chair.
b. Andy and Billy. \(\rightarrow\) Only Andy and Billy can chair.

Nevertheless, if a \(\Diamond\)-question is plural-marked as in \textbf{[4b]} (viz., the \textit{wh}-complement is morphologically plural-marked), mention-some readings seem to become unavailable.

a. Where can we get coffee? \(\approx\) ‘Tell me one/every place where we can get coffee.’
b. At which places can we get coffee? \(\approx\) ‘Tell me #one/every place where we can get coffee.’

This paper investigates the following two issues on the availability of mention-some. First, what licenses mention-some readings in number-neutral \(\Diamond\)-questions? Second, what blocks mention-some readings in plural-marked \(\Diamond\)-questions, and are there exceptions?

2. Competing views on the distribution of mention-some

\textbf{Xiang (2016b: ch. 2)} classifies approaches to mention-some into three groups:

a. \textit{Pragmatic approaches} \textbf{(Groenendijk & Stokhof 1984; van Rooij 2004 a.o.)}: At the semantic level, complete answers must be exhaustive. Mention-some answers are special partial answers that are sufficient for the conversational goal behind the question.

b. \textit{Post-structural approaches} \textbf{(Beck & Rullmann 1999; George 2011; ch. 2)}: At the semantic level, all questions are ambiguous between a mention-some and an exhaustive reading, which are derived via different operations on question roots. Their distributions are only restricted by pragmatic factors.

c. \textit{Structural approaches} \textbf{(George 2011: ch. 6; Fox 2013, 2015 Xiang 2016c,b)}: The mention-some/mention-all ambiguity comes from a structural variation within the question nucleus. \(\Diamond\)-questions admit mention-some due to the presence of an existential modal.
This three-way distinction can be broken down into two core issues: (i) whether mention-some readings are derived at the semantic level, and (ii) whether the distribution of mention-some is restricted by any grammatical factors.

(6)  Summary of current lines of approaches on mention-some (Xiang 2016b: ch. 2)

<table>
<thead>
<tr>
<th></th>
<th>Pragmatic</th>
<th>Post-structural</th>
<th>Structural</th>
</tr>
</thead>
<tbody>
<tr>
<td>...semantically licensed?</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>...grammatically restricted?</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Both pragmatic approaches and post-structural approaches pursue a **pragmatic view** on the distribution of mention-some: the distribution of mention-some is purely determined by pragmatic factors. This view predicts that mention-some is available only if a non-exhaustive answer suffices for the conversational goal of the question (Groenendijk & Stokhof 1984, van Rooij 2004 a.o.). For example, for the $\lozenge$-question in (2), if the conversational goal is to find a place to get some coffee, then naming one accessible coffee place is sufficient, which therefore licenses the mention-some reading. If the conversational goal is to investigate the local coffee market, then (2) admits only a mention-all reading. Likewise, a non-$\lozenge$-question like “who knows Python?” may take a mention-some reading if the goal is just to find someone to help with a Python problem, and it takes a mention-all reading if the goal is to know the programming skills of a group of candidates.

From the observation that mention-some is not restricted to matrix questions, but can also be observed with embedded $\lozenge$-wh-constructions (as in the indirect questions *John knows where we can get coffee*, and the free relative in *John went to where we can get coffee*), post-structural approaches argue that a mention-some reading must be available at the semantic level. Hence, post-structural approaches are commonly called “semantic approaches” in the literature. Nevertheless, post-structural approaches predict no grammatical constraints on the distribution of mention-some. For example, Beck & Rullmann (1999) assume that the root denotation of a question is unambiguously a Hamblin-Karttunen intension, and that a question takes mention-some iff the employed answerhood operator has an existential quantificational force.

(7)  a.  $\text{ANS}_1(Q)(w) = \bigcap \{ p : Q(w)(p) \land p(w) \}$  (for mention-all)

    b.  $\text{ANS}_3(Q)(w) = \lambda P_{s,\text{stt}}. \exists p [ P(w)(p) \land Q(w)(p) \land p(w) ]$  (for mention-some)

Since post-structural approaches derive mention-some through generic operations that can apply to any question (e.g., employing the $\text{ANS}_3$-operator), they predict a systematic mention-some/mention-all ambiguity, which can only be resolved by pragmatic considerations.

Contrary to pragmatic approaches and post-structural approaches, structural approaches pursue a **grammatical view** on the distribution of mention-some: the distribution of mention-some is primarily grammatically restricted. This view relies on two major steps. First, Fox (2013) weakens the notion of completeness. Contra the traditional view that only the strongest true answer is complete (Dayal 1996), Fox proposes that a true answer is complete as long as it is max-informative, that is: it is a member of the Hamblin set and is not asymmetrically entailed by any true proposition in this set. This move allows non-exhaustive answers to be complete answers, when a question has multiple complete true
answers. A question takes a mention-some reading iff it can have multiple max-informative true answers.

\[ \text{ANS}_{\text{fox}}(Q)(w) = \{ p : w \in p \in Q \land \forall q[w \in p \rightarrow p \not\subset q] \} \]

Second, structural approaches ascribe the mention-some/mention-all ambiguity to minimal structural variations within the question nucleus (i.e., the IP part of the logical form of a question). The actual derivations are not crucial for this paper. The core consequence is that, for structural approaches, the form of the question nucleus (especially whether a possibility modal is present) is the primary factor affecting the distribution of mention-some.

To decide between the two competing views regarding the distribution of mention-some, we can test the availability of mention-some in questions with and without a possibility modal under the same “mention-some goal” (i.e., a conversational goal that can be sufficed by a mention-some answer). The pragmatic view predicts that under the same mention-some goal, \( \Diamond \)-questions and non-\( \Diamond \)-questions are equally likely to take mention-some readings. In contrast, the grammatical view predicts that under the same mention-some goal, \( \Diamond \)-questions are more likely to take mention-some than non-\( \Diamond \)-questions.

3. Mention-some in plural-marked \( \Diamond \)-questions

What blocks mention-some readings in (at least some of the) plural-marked \( \Diamond \)-questions? Comorovski (2006: 39) claims that mention-some is unavailable if the \( \text{wh} \)-phrase is D-linked (i.e., the \( \text{wh} \)-phrase is of a complex form “which NP”, which typically requires a contextually salient domain of quantification, see Pesetsky 1987). This claim seems to be consistent with the following data:

\[ a. \quad \text{Where can we get coffee?} \quad (\text{OK mention-some, OK mention-all}) \]
\[ b. \quad \text{At which places can we get coffee?} \quad (#\text{mention-some, OK mention-all}) \]
\[ c. \quad \text{At which place can we get coffee?} \quad (\text{mention-some = mention-all}) \]
\[ \Leftrightarrow \text{There is only one place where we can get coffee.} \]

\[ 1\]For example, Fox (2013) assumes that the \( \text{wh} \)-trace \( X \) has a covert phrase-mate \( \text{EACH} \), and that the scope ambiguity of \([X \text{EACH}]\) relative to the possibility modal is responsible to the mention-some/mention-all ambiguity. Xiang (2016b) provides two ways to capture the mention-some/mention-all ambiguity. One way, which captures the contrast between the mention-some answer\[2a\] and the conjunctive mention-all answer\[2b\], is based on the scope ambiguity of the second-order \( \text{wh} \)-trace (interpreted as a variable of type \( \langle et, t \rangle \)) relative to the possibility modal; the other way, which captures the contrast between the mention-some answer\[2a\] and the disjunctive mention-all answer\[2c\], is based on the optional presence of a covert operator that has a meaning akin to the Mandarin particle \( \text{dou} \) above the possibility modal.

\[ 2\]As pointed to us by Matthijs Westera (p.c.), \( \Diamond \)-questions could be used more frequently in conversations with a mention-some goal, and this may bias them towards a mention-some reading when the context does not fully specify the conversational goal. Such a hypothesis would lead pragmatic approaches to make predictions more in line with the grammatical view. We tried to set a clear goal in the experiments we present below, but this objection seems very hard to fully dismiss.
Nevertheless, the unavailability of mention-some in (9c) is due to an independent reason: the singular wh-complement has a uniqueness presupposition (Srivastav 1991), which collapses mention-some and mention-all.

Alternatively, Xiang (2015), Xiang (2016b: ch. 3) and Dayal (2017: ch. 3) argue that the blocking of mention-some in plural-marked questions comes from the fact that a plural-marked question expects an answer that names a non-atomic individual. In particular, Xiang analyzes this plurality requirement as an anti-presupposition: first, in spirit of Maximize Presupposition (Heim 1991), a plural-marked question anti-presupposes that the corresponding singular-marked question is undefined (a la Sauerland et al. 2005); second, in spirit of question-answer congruence, a proper answer of a plural question needs to entail the anti-presupposition of this question. The plurality-based view predicts that the availability of mention-some for a plural-marked wh-question depends on the type of predicate it combines with, as illustrated by the following minimal pair:

\[(10) \quad \text{(Context: The committee should have only one chair but multiple members.)} \]

\[\begin{align*}
\text{a.} & \quad \text{Which professors can chair the committee? (\#mention-some, OK mention-all)} \\
& \quad \text{# Andy can chair the committee.} \\
\text{b.} & \quad \text{Which professors can form the committee? (OK mention-some, OK mention-all)} \\
& \quad \text{Andy and Billy can form the committee.}
\end{align*}\]

Provided co-chair is disallowed, the question (10a) rejects mention-some because each expected mention-some answer names exactly one atomic individual (e.g., Andy) who can be the committee chair. In contrast, with the collective predicate can form the committee, which can take non-atomic arguments, the question (10b) admits a mention-some reading because a mention-some answer names the sum of a group of individuals (e.g., Andy ⊕ Billy) who together can form the committee, thus satisfying the plurality requirement.

4. **Experiment 1**

4.1 **Goal and design**

The goal of this first experiment was to test the effects of the following two factors on the availability of mention-some readings for questions with collective predicates: presence/absence of a possibility modal and the from of the wh-item (bare wh-words like who vs. complex phrases of the form “which-NP_{PL}”). We used indirect questions because they allowed us to test the exhaustiveness of a question using a simple truth-value judgment task (Cremers & Chemla 2016; Phillips & George 2016). We used the question-embedding verb remember, which sounded more natural than know in the tested context. The background story ensured that the presupposition of past knowledge from remember was satisfied. Moreover, it set up a clear "mention-some" conversational goal, and ensured that the goal be the same for all questions, minimizing the role of pragmatic factors. Questions with collective predicates (such as which children can lead the dance) are crucial for this study, because they can have mention-some answers naming a sum of individuals.
Our experimental settings required participants to compare the facts regarding a question to an agent’s knowledge of these facts. Concretely, this means that they would have to compare two sets of sets. The main challenge was therefore to make it easy for participants to quickly grasp such complex situations and to evaluate the truth of a knowledge-attribution. Our solution to this problem can be conceptualized as follows: a background story established that a pair of children can lead the dance iff they have a piece of clothing in common. This way, we can represent the extension of the collective predicate can lead the dance in a single picture, without having to list all the possible pairs of children who can lead the dance together. For example, the picture in (11) represents in a concise manner a situation in which two pairs of children can lead the dance, namely Ann with Chloe (who wear the same hat) and Bill with Chloe (who wear the same scarf). To prevent a superficial strategy which would have consisted in visually comparing the depictions of reality and of the memory of the agent, Mary’s knowledge state was described in words rather than using the same type of picture.

(11) Example of a target item in Experiment 1

<table>
<thead>
<tr>
<th>How children are dressed up:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann</td>
</tr>
<tr>
<td>Mary’s memory:</td>
</tr>
<tr>
<td>Bill and Chloe wear the same scarf, Chloe wears a hat.</td>
</tr>
<tr>
<td>Therefore Bill and Chloe can lead the dance.</td>
</tr>
</tbody>
</table>

“Mary remembers which children can lead the dance”

4.2 Methods and Material

4.2.1 Instructions

The instructions introduced the following background story:

A party is organized at the school. It will begin with a bal, and Mary is in charge of choosing two children who will lead the dance. The director of the school was very clear on one point: the children leading the dance should have an accessory in common. There is however one exception: if a child is wearing a jester hat, he or she can lead the dance alone.
The sentence about jester hats played no role for target items. It was used to build control items with atomic mention-some answers (see below). The instructions also presented two example items (one clearly true, and one clearly false).

4.2.2 Stimuli

All items involved four children, Ann, Bill, Chloe and Diana, who were always presented in the same order. Each child wore one, two, or no accessory, chosen from 3 simple hats, a scarf, and a bowtie. For target items, one randomly chosen child wore 2 accessories (one headwear and one neckwear), another child had the same headwear but no neckwear, a third child the same neckwear but no headwear, and the last child wore no accessory at all. Hence, in the target items, there were always two pairs of children sharing an accessory, and who could thus lead the dance together. Which child wore the two accessories, headwear only, and neckwear only, was pseudo-randomized.

We manipulated three factors. The first two determined the form of the embedded question in the target sentence: \textsc{whtype}: neutral (who) versus plural (which children) and \textsc{modal}: modalized (can lead the dance) versus plain (have an accessory in common). The third factor, \textsc{situation}, specified Mary’s memory as to who wore what: \textbf{T} (True: Mary correctly remembered who wore what), \textbf{UA} (Under-affirming: Mary’s memory was incomplete; Mary forgot one of the accessories worn by the children, and hence she missed one possible pair of dance leaders), \textbf{OA} (Over-affirming: Mary thought that one of the children wore an additional accessory, which was common with another child. Therefore, she falsely believed that one additional pair of children could lead the dance together), and \textbf{F} (False: all of Mary’s memories were off). The different situations make different readings of the target sentence true or false, as summarized in (12):

\begin{equation}
\begin{array}{|c|c|c|c|c|}
\hline
\text{reading type} & \text{UA} & \text{OA} & \text{F} \\
\hline
\text{mention-all} & 1 & 0 & 0 & 0 \\
\hline
\text{mention-some} & 1 & 1 & 0 & 0 \\
\text{FA-sensitive} & 1 & 1 & 1 & 0 \\
\text{FA-insensitive} & 1 & 1 & 1 & 0 \\
\hline
\end{array}
\end{equation}

Crucially for this study, the \textbf{UA} situation distinguishes between exhaustive/mention-all and mention-some. Mary’s memory in the \textbf{UA} situation was incomplete but involved no false belief. The \textbf{OA} situation distinguishes between false answer (FA-)sensitive readings and FA-insensitive readings. For example, for mention-all readings, the \textbf{OA} situation distinguishes between strongly/intermediately exhaustive (Groenendijk & Stokhof 1984) and weakly exhaustive (Karttunen 1977) readings, among which the former are FA-sensitive and require the belief holder to have no false belief relevant to the embedded question. This sub-categorization also applies to mention-some readings (George 2013, a.o.).

This design led to \(2 \times 2 \times 4\) conditions. \textbf{T} and \textbf{F} situations were repeated three times each, while \textbf{UA} and \textbf{OA} were repeated six times each, for a total of 72 target items.
Right after the instructions, participants were presented with four training items. In case of mistakes, participants received detailed feedback. The feedback screen was displayed for 10s in order to let participants read it carefully, as well as to encourage them to pay more attention. The transition to the experimental phase was not indicated.

In addition to the 72 targets, the experimental phase contained 18 “Jester hat” controls which aimed to test for mention-some readings with atomic responses. When a child wore a jester hat, he or she was allowed to lead the dance alone. This allowed us to build situations in which the question “Who can lead the dance?” had mention-some answers based on atomic individuals. If the plurality-based view (Xiang 2015, 2016b; Dayal 2017) is correct, “Which children can lead the dance?” should reject mention-some in these situations. The jester-hat controls were built on a 3 (WhType-Modal combination) $\times$ 2 (Situation) sub-design. Each combination of conditions was repeated three times, for 18 items in total. Since we were mainly interested in the interaction between can and which-NP_{pl}, we dropped the neutral-plain combination but kept the other three WhType-Modal combinations. To focus on the distribution of mention-some, we tested only two situations: T and UA. All pictures contained two children wearing a jester hat, so that there were two true atomic mention-some answers (each child can lead the dance) and one true non-atomic mention-some answer (the two children can lead the dance together). The plain predicate was wear a jester hat, and the modalized predicate was again can lead the dance.

Finally, we included fillers with declarative complements. Half of the fillers were built with a modalized predicate, and half with a plain predicate. Each type of complement appeared 6 times as a true filler, and 6 times as a false filler, for a total of 24 items.

### 4.2.3 Participants

We recruited 30 participants on Amazon’s Mechanical Turk, and paid $1.80 for each. All reported being native speakers of English. Their age ranged from 19 to 73 year old (mean: 38). Five participants were removed from the analyses because their error rate was one standard deviation above the mean (threshold: 33.5%).

### 4.3 Results and discussion

The results on target items are presented in (13a). The fastest and slowest 1% responses were dropped. The data was analyzed with mixed-effects logit models. The random effects structure for all models was built following the recommendations of Bates et al. (2015). The significance of fixed effects was assessed by comparison with models without the corresponding fixed effects (Levy 2014). We fitted a model on UA targets with WhType, MODAL and their interaction as fixed effects. We observed a significant effect of MODAL ($\chi^2(1) = 41, p < .001$), but no effect of WhType ($\chi^2(1) = 0.61, p = 0.44$) and, crucially, no interaction ($\chi^2(1) = 0.60, p = 0.44$). This shows that mention-some readings are more readily available when the embedded question contains a possibility modal, and that who and which-NP_{pl} had very similar behaviors with respect to mention-some.
Nevertheless, if the predicate that the wh-phrase combines with is distributive, we expect a strong difference between who and which-NP\textsubscript{pl}. In this case, mention-some answers should be based on atomic individuals, and mention-some readings should be observed with who-questions but not with which-questions.

We now turn to the results of the ‘Jester hat’ controls, presented in (13b). A model fitted on the three UA-conditions showed no significant effect of \textsc{modal} ($\chi^2(1) = .07, p = 0.79$) or \textsc{wh-type} ($\chi^2(1) = 0.89, p = 0.34$). If anything, the model with only \textsc{wh-type} as a predictor was slightly better than the model with only \textsc{modal} ($\Delta AIC = 0.82, \Delta BIC = 0.83$), and better than the model with both \textsc{modal} and \textsc{wh-type} as predictors ($\Delta AIC = 1.9, \Delta BIC = 5.4$). This effect seems to suggest the following: when atomic mention-some answers are possible, \textsc{wh-type} becomes an important predictor of mention-some readings, in that modalized which-questions are suddenly closer to their non-modalized counterparts than the modalized who-questions. Nevertheless, this effect was far from significant.

In short, we observed that it’s the presence of a possibility modal in the question that primarily licenses mention-some, and that the type of wh-phrase does not have a significant effect. The first result is a strong argument for the grammatical approaches (George 2011; Fox 2013; Xiang 2016b). The second is in line with Xiang (2015, 2016b) and Dayal (2017), who predict that plural questions should admit mention-some readings when the plurality requirement is satisfied.
Nevertheless, even in jester-hat controls, where mention-some answers involve atomic individuals and do not satisfy Xiang and Dayal’s plurality requirement, we still found no clear contrast between who and which-NP\textsubscript{PL}. This suggests a different explanation for the lack of such contrast in target conditions. Since the task was difficult, it could just be that the participants focused on the presence/absence of a modal, a contrast which is both more visible (it affect all the end of the sentence) and relevant (since modals have long been argued to play a crucial role in the availability of mention-some readings).

5. Experiment 2

5.1 Goal and design

The goal of Experiment 2 was to distinguish between two interpretations of the results from Experiment 1. We made the \textit{WHTYPE} factor between-participants, so that the type of the \textit{wh}-phrase remained the same across trials.

This move also allowed us to reduce the number of conditions each participant would be exposed to. We added another situation called over-denying. In this situation, Mary’s memory would be not only incomplete (like under-affirming) but also involve an explicit negative false belief. Xiang’s (2016a) experiments suggest that mention-some and mention-all knowledge attributions differed in their sensitivity to false beliefs: mention-all is more tolerant of over-affirming, while mention-some is more tolerant of over-denying. We wanted to see if this result could be replicated in a different experimental context.

5.2 Methods and Material

5.2.1 Differences with Experiment 1

Beyond making \textit{WHTYPE} between-participants, the main novelty of Experiment 2 was the addition of an OD (over-denying) level to the \textit{SITUATION} factor. The OD situations was nearly identical to the UA situation presented in (11), except that the description of Mary’s memory involved an explicit exhaustication: “Bill and Chloe wear the same scarf, only Chloe wears a hat. Therefore only Bill and Chloe can lead the dance.” (underlining added for clarity, but not present in actual experimental items). \textit{SITUATION} thus had 5 levels now: T, UA, OA, OD, and F. The other factors were, as in Experiment 1, \textit{MODAL} (2 levels, within-subject), and \textit{WHTYPE} (2 levels, now between-subjects). Each T or F combination was repeated 3 times, while the other combinations were each repeated 6 times. Each participant thus saw \((2 \times 3 + 3 \times 6) \times 2 = 48\) target items.

The neutral version presented the 6 neutral-modalized jester-hat control items, while the plural version presented the 6 plural-modalized and the 6 plural-plain items. Due to a programming error, the UA jester-hat items were mistakenly phrased as OD items.

The rest of the experiment (including the instructions and the declarative controls) was identical to Experiment 1. In total, participants saw 78 items in the neutral version or 84 items in the plural version.
5.2.2 Participants

We recruited 48 participants (24 for the neutral version, 24 for the plural version), who were paid $1.90 each. All reported being native speakers of English. Their ages ranged from 21 to 58 year old (mean: 33). Twelve participants were removed from the analyses because their error rate was one standard deviation above the mean (threshold: 38.5%).

5.3 Results and Discussion

(14) Distribution of participants’ average responses on target items in Experiment 2:

The results on target items in Experiment 2 are presented in (14). We first ran the same statistical analysis as in Experiment 1, focusing on UA targets. Again, there was a clear effect of MODAL ($\chi^2(1) = 6.1, p = 0.014$) but no effect of WH TYPE ($\chi^2(1) = 1.7, p = 0.20$). Interestingly, we now observed a significant interaction between these two factors ($\chi^2(1) = 6.1, p = 0.013$): the modal had a clear effect on the neutral items but not on the plural items. In particular, modalized who-questions were even more likely to take mention-some readings than they were in Experiment 1, while which children-questions overall received lower acceptance rates in both UA conditions.

Crucially, the availability of mention-some readings with who-questions and their dependency on the presence of a modal were even clearer in Experiment 2, while which-NP_{PL}-
questions received the same average acceptance rates in the UA conditions, but without any influence of the presence of the modal. This difference between *who* and *which*-NP<sub>pl</sub> confirms our worry that their near-identical behavior in Experiment 1 was due to a confusion on the part of participants, who may have focused more on MODAL and dismissed WHICH.<sup>3</sup>

We then ran a model on the OA and OD neutral conditions, after selecting the participants who accepted the UA condition at least 50% of the time (i.e., who had a mention-some reading). We observed a strong effect of MODAL ($\chi^2(1) = 13, p < .001$), but no main effect of SITUATION ($\chi^2(1) = 0.5, p = 0.49$) or interaction of these two factors ($\chi^2(1) = 1.9, p = 0.17$). This suggests that participants were simply more tolerant of false beliefs in modalized questions, but we did not replicate Xiang’s (2016a) findings on the asymmetry between over-affirming and over-denying. Nevertheless, this statistical analysis is clearly underpowered, since only 10 participants satisfied all criteria (i.e., not being excluded on the basis of errors, falling in the neutral condition, and accepting UA situations).

In summary, Experiment 2 replicated the results of Experiment 1 with *who*-questions, but we observed less clear evidence for mention-some readings with *which*-NP<sub>pl</sub>-questions this time. In particular, unlike the case of *who*-questions, MODAL seemed to play little role in the case of *which*-NP<sub>pl</sub>-questions.

### 6. Conclusion

First of all, our results validate the role of the possibility modal in licensing the mention-some reading of *who*-questions. This conclusion was the clearest one from both experiments, and it is strong evidence in favor of the grammatical view against the pragmatic view on the distribution of mention-some. Indeed, while the context remained constant, we saw that the presence of a possibility modal was necessary to license the mention-some reading of the embedded *who*-question.

By contrast, the availability of mention-some in *which*-NP<sub>pl</sub>-questions was less clear. In Experiment 1, we observed no difference between modalized *who*-questions and modalized *which*-NP<sub>pl</sub>-questions: in presence of a possibility modal, both types of questions gave rise to mention-some readings at the same rate. This result, on the one hand, seems to be compatible with the prediction of the analyses by Xiang (2015, 2016b) and (Dayal 2017: ch. 3): plural-marked *which*-questions can give rise to mention-some readings, provided there are mention-some answers naming non-atomic individuals. On the other hand, the lack of contrast in jester-hat controls suggested an alternative explanation: participants may have dismissed the variations of WHICH and treated all questions as *who*-questions. In Experiment 2, controlling for this potential confound, we observed much lower acceptance rates for UA situations in *which*-NP<sub>pl</sub>-questions than in modalized *who*-questions. This observation suggests that *which*-NP<sub>pl</sub>-questions are in fact less likely to take mention-some readings than modalized *who*-questions. We also observed that the acceptance rates for UA situations in *which*-NP<sub>pl</sub>-questions were independent from whether the questions were modalized. This result supports Comorovski’s (2006) generalization (that D-linked

---

<sup>3</sup>Note that we were not able to test situations where atomic responses are available, because of the programming error on jester-hat controls. However, given that participants made a clear distinction between the two types of WH-PHRASES on target items, this control was not crucial anymore.
wh-phrases unlicense mention-some readings), and would call for an independent mechanism of pragmatic tolerance for partial answers when they satisfy the conversation goals in order to account for residual acceptance of the UA conditions with which-NP\textsubscript{PL}. Such a mechanism, which would not be sensitive to the presence of a modal or other structural factors, is needed anyway to explain the “who knows Python?” cases.

An other possibility, which potentially rescues the view of Xiang (2015, 2016b) and Dayal (2017), is that the predicate lead the dance can be interpreted either collectively or distributively, and that using a complex which-phrase favors the distributive reading:

\begin{equation}
\text{(15) Which children can lead the dance?}
\end{equation}

\begin{enumerate}
\item \textit{Collective reading}\n\begin{quote}
‘Which (groups of) children can lead the dance together?’
\end{quote}
\item \textit{Distributive reading}\n\begin{quote}
‘Which children can be part of a dance-leading pair?’
\end{quote}
\end{enumerate}

Under the distributive reading, mention-some answers each name only one atomic individual and thus are excluded by the plurality requirement.\footnote{In \textit{(15b)} the distributive operator comes within the embedded predicate lead the dance and thus scopes below the possibility modal, and the mention-some reading of this form is blocked by the plurality requirement. An alternative form of distributivity proposed by Williams (2000) might also account for the unavailability of mention-some. To explain quantificational variability effects in questions with collective predicates (e.g., \textit{John knows for the most part which students formed the bassoon quintet}), Williams proposes a complex lexical entry for the determiner which that has a built-in distributive operator. If we follow Williams’s idea, we predict yet a different distributive logical form for \textit{(15)} since the distributive operator would be part of which (paraphrased as ‘which children are part of groups who can lead the dance?’), it would have to take a wide scope, yielding only a mention-all reading according to existing grammatical theories.}

\textbf{References}


Fox, Danny. 2013. Mention-some readings of questions. \textit{MIT seminar notes}.
Fox, Danny. 2015. Mention some, reconstruction, and the notion of answerhood. Handout at Experimental and crosslinguistic evidence for the distinction between implicatures and presuppositions (ImPres).


